Express Mail Label No. EV301223183US
Docket No.: 393032042000
(PATENT)

IN THE UNITED STATES PATENT AND TRADEMARK OFFICE

In re Patent Application of:
Takahiro KAWASHIMA

Application No.: Not Yet Assigned

Filed: Concurrently Herewith

Art Unit: Not Yet Assigned

For: INTERCHANGE FORMAT OF VOICE DATA IN MUSIC FILE

Examiner: Not Yet Assigned

CLAIM FOR PRIORITY AND SUBMISSION OF DOCUMENT

MS Patent Application
Commissioner for Patents
P.O. Box 1450
Alexandria, VA 22313-1450

Dear Sir:

Applicant hereby claims priority under 35 U.S.C. 119 based on the following prior foreign application filed in the following foreign country on the date indicated:

Country	Application No.	Date
Japan	2002-335233	November 19, 2002

In support of this claim, a certified copy of the said original foreign application is filed herewith.

Dated: November 17, 2003

Respectfully submitted,

By _____

David L. Fehrman
Registration No.: 28,600
MORRISON & FOERSTER LLP
555 West Fifth Street, Suite 3500
Los Angeles, California 90013
(213) 892-5587

日 本 国 特 許 庁
JAPAN PATENT OFFICE

別紙添付の書類に記載されている事項は下記の出願書類に記載されている事項と同一であることを証明する。

This is to certify that the annexed is a true copy of the following application as filed with this Office.

出 願 年 月 日
Date of Application: ２００２年１１月１９日

出 願 番 号
Application Number: 特願２００２－３３５２３３

[ST. 10/C]: [ＪＰ２００２－３３５２３３]

出 願 人
Applicant(s): ヤマハ株式会社

２００３年 ９月２２日

特許庁長官
Commissioner,
Japan Patent Office 今井康夫

出証番号 出証特２００３－３０７７６４１

【書類名】　特許願
【整理番号】　YC30621
【提出日】　平成14年11月19日
【あて先】　特許庁長官殿
【国際特許分類】　G10L　3/00
【発明者】
　【住所又は居所】　静岡県浜松市中沢町１０番１号　ヤマハ株式会社内
　【氏名】　川嶋　隆宏
【特許出願人】
　【識別番号】　000004075
　【氏名又は名称】　ヤマハ株式会社
【代理人】
　【識別番号】　100102635
　【弁理士】
　【氏名又は名称】　浅見　保男
【選任した代理人】
　【識別番号】　100106459
　【弁理士】
　【氏名又は名称】　高橋　英生
【選任した代理人】
　【識別番号】　100105500
　【弁理士】
　【氏名又は名称】　武山　吉孝
【選任した代理人】
　【識別番号】　100103735
　【弁理士】
　【氏名又は名称】　鈴木　隆盛

【手数料の表示】

【予納台帳番号】　037338

【納付金額】　21,000円

【提出物件の目録】

【物件名】　明細書　1

【物件名】　図面　1

【物件名】　要約書　1

【包括委任状番号】　9808721

【包括委任状番号】　0106838

【プルーフの要否】　要

【書類名】　明細書

【発明の名称】　シーケンスデータのデータ交換フォーマット、音声再生装置及びサーバー装置

【特許請求の範囲】

【請求項１】　音源を用いて音声を再生させる音声再生シーケンスデータのデータ交換フォーマットであって、

管理用の情報を含むコンテンツデータチャンクと、音声再生シーケンスデータを含むトラックチャンクを有し、

前記音声再生シーケンスデータは、音声の再生を指示する音声再生イベントと、その音声再生イベントを実行するタイミングを先行する音声再生イベントからの経過時間により指定するデュレーションとの組が時間順に配置されているものであることを特徴とするシーケンスデータのデータ交換フォーマット。

【請求項２】　楽曲を再生させるための楽曲シーケンスデータと音声を再生させるための音声再生シーケンスデータとを含むシーケンスデータのデータ交換フォーマットであって、

前記楽曲シーケンスデータは、演奏イベントデータとその演奏イベントを実行するタイミングを先行する演奏イベントからの経過時間により指定するデュレーションデータとの組が時間順に配置されたものであり、

前記音声再生シーケンスデータは、音声の再生を指示する音声再生イベントデータとその音声再生イベントを実行するタイミングを先行する音声再生イベントからの経過時間により指定するデュレーションデータとの組が時間順に配置されたものであり、

前記楽曲シーケンスデータと前記音声再生シーケンスデータの再生を同時に開始させることにより、当該楽曲と当該音声とを同じ時間軸上で再生することができるようになされていることを特徴とするシーケンスデータのデータ交換フォーマット。

【請求項３】　前記楽曲シーケンスデータと前記音声再生シーケンスデータはそれぞれ異なるチャンクに含まれていることを特徴とする請求項２記載のシーケンスデータのデータ交換フォーマット。

【請求項４】　前記音声再生イベントデータは、（１）合成される音声の読みを示すテキスト情報と音声表現を指定する韻律記号とからなるテキスト記述型の情報、（２）合成される音声を示す音素情報と韻律制御情報とからなる音素記述型の情報、又は、（３）再生される音声を示すフレーム時間毎のフォルマント制御情報からなるフォルマントフレーム記述型の情報、の再生を指示するデータであることを特徴とする請求項１乃至３のいずれかに記載のシーケンスデータのデータ交換フォーマット。

【請求項５】　音声の再生を指示する音声再生イベントデータと、その音声再生イベントを実行するタイミングを先行する音声再生イベントからの経過時間により指定するデュレーションデータとの組により構成されているシーケンスデータに基づいて音声を再生する音声再生装置であって、

テキスト情報及び韻律記号と音素情報及び韻律制御情報との対応を記録した辞書を参照して、テキスト情報及び韻律記号に対応する音素情報及び韻律制御情報を取得する第１の手段と、

音素情報及び韻律制御情報とフォルマント制御情報との対応を記録した辞書を参照して、音素情報及び韻律制御情報に対応するフォルマント制御情報を取得する第２の手段と、

フォルマント制御情報をフレーム毎に音源部に出力する出力手段とを有し、

前記音声再生イベントデータが合成される音声の読みを示すテキスト情報と韻律記号とからなるテキスト記述型の情報の再生を指示するデータであるときは、前記第１の手段、前記第２の手段及び前記出力手段を用いて当該音声を再生し、

前記音声再生イベントデータが合成される音声に対応する音素情報と韻律制御情報を含む音素記述型の情報の再生を指示するデータであるときは、前記第２の手段と前記出力手段を用いて当該音声を再生し、

前記音声再生イベントデータがフレーム毎のフォルマント制御情報の再生を指示するデータであるときは、前記出力手段を用いて当該音声を再生する

ことを特徴とする音声再生装置。

【請求項６】　音声の再生を指示する音声再生イベントデータと、その音声再生イベントを実行するタイミングを先行する音声再生イベントからの経過時間

により指定するデュレーションデータとの組により構成されているシーケンスデータに基づいて音声を再生する音声再生装置であって、

音素情報及び韻律制御情報とフォルマント制御情報との対応を記憶した辞書を参照して、音素情報及び韻律制御情報に対応するフォルマント制御情報を取得する手段と、

フォルマント制御情報をフレーム毎に音源部に出力する出力手段とを有し、

前記音声再生イベントデータが合成される音声に対応する音素情報と韻律制御情報を含む音素記述型の情報の再生を指示するデータであるときは、前記取得する手段と前記出力手段とを用いて当該音声を再生し、

前記音声再生イベントデータがフレーム毎のフォルマント制御情報の再生を指示するデータであるときは、前記出力手段を用いて当該音声を再生する

ことを特徴とする音声再生装置。

【請求項７】　音声の再生を指示する音声再生イベントデータと、その音声再生イベントを実行するタイミングを先行する音声再生イベントからの経過時間により指定するデュレーションデータとの組により構成されているシーケンスデータに基づいて音声を再生する音声再生装置であって、

フォルマント制御情報をフレーム毎に音源部に出力する出力手段を有し、

前記音声再生イベントデータがフレーム毎のフォルマント制御情報の再生を指示するデータであるときに、前記出力手段を用いて当該音声を再生することを特徴とする音声再生装置。

【請求項８】　シーケンスデータを蓄積し、クライアント装置からの要求に応じて対応するシーケンスデータを配信するサーバー装置であって、

前記シーケンスデータは、

楽曲シーケンスデータを含むチャンクと音声再生シーケンスデータを含むチャンクを有し、

前記楽曲シーケンスデータは、演奏イベントデータとその演奏イベントを実行するタイミングを先行する演奏イベントからの経過時間により指定するデュレーションデータとの組が時間順に配置されたものであり、

前記音声再生シーケンスデータは、音声の再生を指示する音声再生イベントデ

ータとその音声再生イベントを実行するタイミングを先行する音声再生イベントからの経過時間により指定するデュレーションデータとの組が時間順に配置されたものである

ことを特徴とするサーバー装置。

【請求項９】　前記音声再生イベントデータは、（１）合成される音声の読みを示すテキスト情報と韻律記号とからなるテキスト記述型の情報、（２）合成される音声を示す音素情報と韻律制御情報とからなる音素記述型の情報、又は、（３）再生される音声を示すフレーム時間毎のフォルマント制御情報からなるフォルマントフレーム記述型の情報、の再生を指示するデータであり、

クライアント装置からの要求に応じて、前記（１）～（３）のうちのいずれかの音声再生イベントデータを含む音声再生シーケンスデータを選択して、当該クライアント装置に配信することを特徴とする請求項８記載のサーバー装置。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】

本発明は、シーケンスデータのデータ交換フォーマット、音声再生装置及びサーバー装置に関する。

【０００２】

【従来の技術】

音源を用いて音楽を表現するためのデータを頒布したり、相互に利用したりするためのデータ交換フォーマットとして、ＳＭＦ（Standard MIDI file format）やＳＭＡＦ（Synthetic Music Mobile Application Format）などが知られている。ＳＭＡＦは、携帯端末などにおいてマルチメディアコンテンツを表現するためのデータフォーマット仕様である（非特許文献１参照）。

【０００３】

図１５を参照しつつＳＭＡＦについて説明する。

この図において、１００はＳＭＡＦファイルであり、チャンクとよばれるデータの塊が基本構造となっている。チャンクは固定長（８バイト）のヘッダ部と任意長のボディ部とからなり、ヘッダ部は、さらに、４バイトのチャンクＩＤと４

バイトのチャンクサイズに分けられる。チャンクＩＤはチャンクの識別子に用い、チャンクサイズはボディ部の長さを示している。ＳＭＡＦファイルは、それ自体及びそれに含まれる各種データも全てチャンク構造となっている。

この図に示すように、ＳＭＡＦファイル１００の中身は、管理用の情報が格納されているコンテンツ・インフォ・チャンク（Contents Info Chunk）１０１と、出力デバイスに対するシーケンスデータを含む１つ以上のトラックチャンク１０２～１０８とからなる。シーケンスデータは出力デバイスに対する制御を時間を追って定義したデータ表現である。１つのＳＭＡＦファイル１００に含まれる全てのシーケンスデータは時刻０で同時に再生を開始するものと定義されており、結果的に全てのシーケンスデータが同期して再生される。

シーケンスデータはイベントとデュレーションの組み合わせで表現される。イベントは、シーケンスデータに対応する出力デバイスに対する制御内容のデータ表現であり、デュレーションは、イベントとイベントとの間の経過時間を表現するデータである。イベントの処理時間は実際には０ではないが、ＳＭＡＦのデータ表現としては０とみなし、時間の流れは全てデュレーションで表わすようにしている。あるイベントを実行する時刻は、そのシーケンスデータの先頭からのデュレーションを積算することで一意に決定することができる。イベントの処理時間は、次のイベントの処理開始時刻に影響しないことが原則である。従って、値が０のデュレーションを挟んで連続したイベントは同時に実行すると解釈される。

【０００４】

ＳＭＡＦでは、前記出力デバイスとして、ＭＩＤＩ（musical instrument digital interface）相当の制御データで発音を行う音源デバイス１１１、ＰＣＭデータの再生を行うＰＣＭ音源デバイス（ＰＣＭデコーダ）１１２、テキストや画像の表示を行うＬＣＤなどの表示デバイス１１３などが定義されている。

トラックチャンクには、定義されている各出力デバイスに対応して、スコアトラックチャンク１０２～１０５、ＰＣＭオーディオトラックチャンク１０６、グラフィックストラックチャンク１０７及びマスタートラックチャンク１０８がある。ここで、マスタートラックチャンクを除くスコアトランクチャンク、ＰＣＭ

オーディオトラックチャンク及びグラフィックストラックチャンクは、それぞれ最大２５６トラックまで記述することが可能である。

図示する例では、スコアトラックチャンク１０２～１０５は音源デバイス１１１を再生するためのシーケンスデータを格納し、ＰＣＭトラックチャンク１０６はＰＣＭ音源デバイス１１２で発音されるADPCMやMP3、TwinVQ等のwaveデータをイベント形式で格納し、グラフィックトラックチャンク１０７は背景画や差込静止画、テキストデータと、それらを表示デバイス１１３で再生するためのシーケンスデータを格納している。また、マスタートラックチャンク１０８にはＳＭＡＦシーケンサ自身を制御するためのシーケンスデータが格納されている。

【０００５】

一方、音声合成の手法として、ＬＰＣなどのフィルタ合成方式や複合正弦波音声合成法などの波形合成方式がよく知られている。複合正弦波音声合成法（ＣＳＭ法）は、複数の正弦波の和により音声信号をモデル化し音声合成を行う方式であり、簡単な合成法でありながら良質な音声を合成することができる。（非特許文献２参照）。

また、音源を用いて音声合成させることにより、歌声を発生させる音声合成装置も提案されている（特許文献１参照）。

【０００６】

【非特許文献１】

ＳＭＡＦ仕様書　Ver. 3.06 ヤマハ株式会社、［平成１４年１０月１８日検索］、インターネット＜URL: http://smaf.yamaha.co.jp＞

【非特許文献２】

嵯峨山茂樹、板倉文忠、「複合正弦波音声合成方式の検討と合成器の試作」、日本音響学会、音声研究会資料、資料番号S80-12(1980-5)、p.93-100、(1980.5.26)

【特許文献１】

特開平９－５０２８７号公報

【０００７】

【発明が解決しようとする課題】

上述のように、ＳＭＡＦは、ＭＩＤＩ相当のデータ（楽曲データ）、ＰＣＭオーディオデータ、テキストや画像の表示データなどの各種シーケンスデータを含み、全シーケンスを時間的に同期して再生することができる。

しかしながら、ＳＭＦやＳＭＡＦには音声（人の声）を表現することについては、定義されていない。

そこで、ＳＭＦなどのＭＩＤＩイベントを拡張して音声を合成することも考えられるが、この場合は、音声部分のみ一括して取り出して音声合成するときに処理が複雑になるという問題点がある。

【０００８】

そこで本発明は、柔軟性があり、かつ、楽曲シーケンスなどと音声再生シーケンスとを同期して再生させることが可能なシーケンスデータのデータ交換フォーマット、該データ交換フォーマットのファイルを再生することができる音声再生装置及び該データ交換フォーマットのデータを配信することができるサーバー装置を提供することを目的としている。

【０００９】

【課題を解決するための手段】

上記目的を達成するために、本発明のシーケンスデータのデータ交換フォーマットは、音源を用いて音声を再生させる音声再生シーケンスデータのデータ交換フォーマットであって、管理用の情報を含むコンテンツデータチャンクと、音声再生シーケンスデータを含むトラックチャンクを有し、前記音声再生シーケンスデータは、音声の再生を指示する音声再生イベントと、その音声再生イベントを実行するタイミングを先行する音声再生イベントからの経過時間により指定するデュレーションとの組が時間順に配置されているものである。

また、本発明の他のシーケンスデータのデータ交換フォーマットは、楽曲を再生させるための楽曲シーケンスデータと音声を再生させるための音声再生シーケンスデータとを含むシーケンスデータのデータ交換フォーマットであって、前記楽曲シーケンスデータは、演奏イベントデータとその演奏イベントを実行するタイミングを先行する演奏イベントからの経過時間により指定するデュレーションデータとの組が時間順に配置されたものであり、前記音声再生シーケンスデータ

は、音声の再生を指示する音声再生イベントデータとその音声再生イベントを実行するタイミングを先行する音声再生イベントからの経過時間により指定するデュレーションデータとの組が時間順に配置されたものであり、前記楽曲シーケンスデータと前記音声再生シーケンスデータの再生を同時に開始させることにより、当該楽曲と当該音声とを同じ時間軸上で再生することができるようになされているものである。

さらに、前記楽曲シーケンスデータと前記音声再生シーケンスデータはそれぞれ異なるチャンクに含まれているものである。

そして、前記音声再生イベントデータは、（１）合成される音声の読みを示すテキスト情報と音声表現を指定する韻律記号とからなるテキスト記述型の情報、（２）合成される音声を示す音素情報と韻律制御情報とからなる音素記述型の情報、又は、（３）再生される音声を示すフレーム時間毎のフォルマント制御情報からなるフォルマントフレーム記述型の情報、の再生を指示するデータとされているものである。

【００１０】

さらにまた、本発明の音声再生装置は、音声の再生を指示する音声再生イベントデータと、その音声再生イベントを実行するタイミングを先行する音声再生イベントからの経過時間により指定するデュレーションデータとの組により構成されているシーケンスデータに基づいて音声を再生する音声再生装置であって、テキスト情報及び韻律記号と音素情報及び韻律制御情報との対応を記録した辞書を参照して、テキスト情報及び韻律記号に対応する音素情報及び韻律制御情報を取得する第１の手段と、音素情報及び韻律制御情報とフォルマント制御情報との対応を記録した辞書を参照して、音素情報及び韻律制御情報に対応するフォルマント制御情報を取得する第２の手段と、フォルマント制御情報をフレーム毎に音源部に出力する出力手段とを有し、前記音声再生イベントデータが合成される音声の読みを示すテキスト情報と韻律記号とからなるテキスト記述型の情報の再生を指示するデータであるときは、前記第１の手段、前記第２の手段及び前記出力手段を用いて当該音声を再生し、前記音声再生イベントデータが合成される音声に対応する音素情報と韻律制御情報を含む音素記述型の情報の再生を指示するデー

タであるときは、前記第２の手段と前記出力手段を用いて当該音声を再生し、前記音声再生イベントデータがフレーム毎のフォルマント制御情報の再生を指示するデータであるときは、前記出力手段を用いて当該音声を再生するものである。

又は、前記第２の手段と前記出力手段を有するもの、もしくは、前記出力手段を有するものである。

【００１１】

さらにまた、本発明のサーバー装置は、シーケンスデータを蓄積し、クライアント装置からの要求に応じて対応するシーケンスデータを配信するサーバー装置であって、前記シーケンスデータは、楽曲シーケンスデータを含むチャンクと音声再生シーケンスデータを含むチャンクを有し、前記楽曲シーケンスデータは、演奏イベントデータとその演奏イベントを実行するタイミングを先行する演奏イベントからの経過時間により指定するデュレーションデータとの組が時間順に配置されたものであり、前記音声再生シーケンスデータは、音声の再生を指示する音声再生イベントデータとその音声再生イベントを実行するタイミングを先行する音声再生イベントからの経過時間により指定するデュレーションデータとの組が時間順に配置されたものとされている。

そして、前記音声再生イベントデータは、（１）合成される音声の読みを示すテキスト情報と韻律記号とからなるテキスト記述型の情報、（２）合成される音声を示す音素情報と韻律制御情報とからなる音素記述型の情報、又は、（３）再生される音声を示すフレーム時間毎のフォルマント制御情報からなるフォルマントフレーム記述型の情報、の再生を指示するデータであり、クライアント装置からの要求に応じて、前記（１）～（３）のうちのいずれかの音声再生イベントデータを含む音声再生シーケンスデータを選択して、当該クライアント装置に配信するようになされている。

【００１２】

【発明の実施の形態】

図１は、本発明における音声再生シーケンスデータのデータ交換フォーマットの一実施の形態を示す図である。この図において、１は本発明のデータ交換フォーマットを有するファイルである。このファイル１は、前述したＳＭＡＦファイ

ルと同様に、チャンク構造を基本としており、ヘッダ部とボディ部とを有する（ファイルチャンク）。

前記ヘッダ部には、ファイルを識別するためのファイルＩＤ（チャンクＩＤ）と後続するボディ部の長さを示すチャンクサイズが含まれている。

ボディ部はチャンク列であり、図示する例では、コンテンツ・インフォ・チャンク（Contents Info Chunk）２、オプショナル・データ・チャンク（Optional Data Chunk）３、及び、音声再生シーケンスデータを含むＨＶ（Human Voice）トラックチャンク４が含まれている。なお、図１には、ＨＶトラックチャンク４として、ＨＶトラックチャンク#00の一つのみが記載されているが、ファイル１中に複数個のＨＶトラックチャンク４を含ませることができる。

また、本発明においては、前記ＨＶトラックチャンク４に含まれる音声再生シーケンスデータとして、３つのフォーマットタイプ（TSeq型、PSeq型、FSeq型）が定義されている。これらについては後述する。

前記コンテンツ・インフォ・チャンク２には、含まれているコンテンツのクラス、種類、著作権情報、ジャンル名、曲名、アーティスト名、作詞/作曲者名などの管理用の情報が格納されている。また、前記著作権情報やジャンル名、曲名、アーティスト名、作詞/作曲者名などの情報を格納するオプショナル・データ・チャンク３を設けても良い。

【００１３】

図１に示した音声再生シーケンスデータのデータ交換フォーマットは、それ単独で音声を再生することができるが、前記ＨＶトラックチャンク４をデータチャンクの一つとして前述したＳＭＡＦファイルに含ませることができる。

図２は、上述したＨＶトラックチャンク４をデータチャンクの一つとして含む本発明のシーケンスデータのデータ交換フォーマットを有するファイルの構造を示す図である。このファイルは、ＳＭＡＦファイルを音声再生シーケンスデータを含むように拡張したものであるということができる。なお、この図において、前記図１５に示したＳＭＡＦファイル１００と同一の要素には同一の番号を付す。

この図に示すように、前述した音声再生シーケンスデータのデータ交換フォー

マットにおけるＨＶトラックチャンク４を、前述したスコアトラックチャンク１０２～１０５、ＰＣＭオーディオトラックチャンク１０６、グラフィックストラックチャンク１０７などと共に、ＳＭＡＦファイル１００中に格納することにより、楽曲の演奏や画像、テキストの表示と同期して音声を再生することが可能となり、例えば、楽音に対し、音源が歌うコンテンツなどを実現することができるようになる。

【００１４】

図３は、前記図２に示した本発明のデータ交換フォーマットのファイルを作成するシステム及び該データ交換フォーマットファイルを利用するシステムの概略構成の一例を示す図である。

この図において、２１はＳＭＦやＳＭＡＦなどの楽曲データファイル、２２は再生される音声に対応するテキストファイル、２３は本発明によるデータ交換フォーマットのファイルを作成するためのデータ・フォーマット制作ツール（オーサリング・ツール）、２４は本発明のデータ交換フォーマットを有するファイルである。

オーサリング・ツール２３は、再生する音声の読みを示す音声合成用テキストファイル２２を入力して、編集作業などを行い、それに対応する音声再生シーケンスデータを作成する。そして、ＳＭＦやＳＭＡＦなどの楽曲データファイル２１に該作成した音声再生シーケンスデータを加えて、本発明のデータ交換フォーマット仕様に基づくファイル（前記図２に示したＨＶトラックチャンクを含むＳＭＡＦファイル）２４を作成する。

【００１５】

作成されたファイル２４は、シーケンスデータに含まれているデュレーションにより規定されるタイミングで音源部２７に制御パラメータを供給するシーケンサ２６と、シーケンサ２６から供給される制御パラメータに基づいて音声を再生出力する音源部２７を有する利用装置２５に転送され、そこで、楽曲などとともに音声が同期して再生されることとなる。

図４は前記音源部２７の概略構成の一例を示す図である。

この図に示した例では、音源部２７は、複数のフォルマント生成部２８と１個

のピッチ生成部２９を有しており、前記シーケンサ２６から出力されるフォルマント制御情報（各フォルマントを生成するためのフォルマント周波数、レベルなどのパラメータ）及びピッチ情報に基づいて各フォルマント生成部２８で対応するフォルマント信号を発生し、これらをミキシング部３０で加算することにより対応する音声合成出力が生成される。なお、各フォルマント生成部２８はフォルマント信号を発生させるためにその元となる基本波形を発生させるが、この基本波形の発生には、例えば、周知のＦＭ音源の波形発生器を利用することができる。

【００１６】

前述のように、本発明においては、前記ＨＶトラックチャンク４に含まれる音声再生シーケンスデータに３つのフォーマットタイプを用意し、これらを任意に選択して用いることができるようにしている。以下、これらについて説明する。

再生する音声を記述するためには、再生する音声に対応する文字情報、言語に依存しない発音情報、音声波形そのものを示す情報など抽象度が異なる各種の段階の記述方法があるが、本発明においては、（ａ）テキスト記述型（TSeq型）、（ｂ）音素記述型（PSeq型）及び（ｃ）フォルマント・フレーム記述型（FSeq型）の３通りのフォーマットタイプを定義している。

【００１７】

まず、図５を参照して、これら３つのフォーマットタイプの相違について説明する。

（ａ）テキスト記述型（TSeq型）

TSeq型は、発音すべき音声をテキスト表記により記述するフォーマットであり、それぞれの言語による文字コード（テキスト情報）とアクセントなどの音声表現を指示する記号（韻律記号）とを含む。このフォーマットのデータはエディタなどを用いて直接作成することができる。再生するときは、図５の（ａ）に示すように、ミドルウェア処理により、該TSeq型のシーケンスデータを、まず、PSeq型に変換し（第１のコンバート処理）、次に、PSeq型をFSeq型に変換（第２のコンバート処理）して、前記音源部２７に出力することとなる。

ここで、TSeq型からPSeq型へ変換する第１のコンバート処理は、言語に依存す

る情報である文字コード（例えば、ひらがなやカタカナなどのテキスト情報）と韻律記号と、それに対応する言語に依存しない発音を示す情報（音素）と韻律を制御するための韻律制御情報を格納した第１の辞書を参照することにより行われ、PSeq型からFSeq型への変換である第２のコンバート処理は、各音素及び韻律制御情報とそれに対応するフォルマント制御情報（各フォルマントを生成するためのフォルマントの周波数、帯域幅、レベルなどのパラメータ）を格納した第２の辞書を参照することにより行われる。

（ｂ）音素記述型（PSeq型）

PSeq型は、ＳＭＦで定義するＭＩＤＩイベントに類似する形式で発音すべき音声に関する情報を記述するものであり、音声記述としては言語依存によらない音素単位をベースとする。図５の（ｂ）に示すように、前記オーサリング・ツールなどを用いて実行されるデータ制作処理においては、まずTSeq型のデータファイルを作成し、これを第１のコンバート処理によりPSeq型に変換する。このPSeq型を再生するときは、ミドルウェア処理として実行される第２のコンバート処理によりPSeq型のデータファイルをFSeq型に変換して、音源部２７に出力する。

（ｃ）フォルマント・フレーム記述型（FSeq型）

FSeq型は、フォルマント制御情報をフレーム・データ列として表現したフォーマットである。図５の（ｃ）に示すように、データ制作処理において、TSeq型→第１のコンバート処理→PSeq型→第２のコンバート処理→FSeq型への変換を行う。また、サンプリングされた波形データから通常の音声分析処理と同様の処理である第３のコンバート処理によりFSeq型のデータを作成することもできる。再生時には、該FSeq型のファイルをそのまま前記音源部に出力して再生することができる。

このように、本発明においては、抽象度の異なる３種類のフォーマットタイプを定義し、個々の場合に応じて、所望のタイプを選択することができるようにしている。また、音声を再生するために実行する前記第１のコンバート処理及び前記第２のコンバート処理をミドルウェア処理として実行させることにより、アプリケーションの負担を軽減することができる。

【００１８】

次に、前記ＨＶトラックチャンク４（図１）の内容について詳細に説明する。

前記図１に示したように、各ＨＶトラックチャンク４には、このＨＶトラックチャンクに含まれている音声再生シーケンスデータが前述した３通りのフォーマットタイプのうちのどのタイプであるかを示すフォーマットタイプ（Format Type）、使用されている言語種別を示す言語タイプ（Language Type）及びタイムベース（Timebase）をそれぞれ指定するデータが記述されている。

フォーマットタイプ（Format Type）の例を表１に示す。

【表１】

フォーマットタイプ	説　明
0x00	TSeq型
0x01	PSeq型
0x02	FSeq型

【００１９】

言語タイプ（Language Type）の例を表２に示す。

【表２】

言語タイプ	説　明
0x00	Shift-JIS
0x02	EUC-KR(KS)

なお、ここでは、日本語（0x00；0xは１６進を表わす。以下、同じ。）と韓国語（0x01）のみを示しているが、中国語、英語などその他の言語についても同様に定義することができる。

【００２０】

タイムベース（Timebase）は、このトラックチャンクに含まれるシーケンスデータチャンク内のデュレーション及びゲートタイムの基準時間を定めるものである。この実施の形態では、20msecとされているが任意の値に設定することができ

る。

【表３】

タイムベース	説　明
0x11	20 msec

【００２１】

前述した３通りのフォーマットタイプのデータの詳細についてさらに説明する。

（ａ）Tseq型（フォーマットタイプ＝0x00）

前述のように、このフォーマットタイプは、テキスト表記によるシーケンス表現（TSeq：text sequence）を用いたフォーマットであり、シーケンスデータチャンク５とｎ個（ｎは１以上の整数）のTSeqデータチャンク（TSeq#00～TSeq#n）６，７，８を含んでいる（図１）。シーケンスデータに含まれる音声再生イベント（ノートオンイベント）でTSeqデータチャンクに含まれるデータの再生を指示する。

【００２２】

(a-1) シーケンスデータチャンク

シーケンスデータチャンクは、ＳＭＡＦにおけるシーケンスデータチャンクと同様に、デュレーションとイベントの組み合わせを時間順に配置したシーケンスデータを含む。図６の（ａ）はシーケンスデータの構成を示す図である。ここで、デュレーションは、イベントとイベントの間の時間を示している。先頭のデュレーション（Duration 1）は、時刻０からの経過時間を示している。図６の（ｂ）は、イベントがノートメッセージである場合に、デュレーションとノートメッセージに含まれるゲートタイムの関係を示す図である。この図に示すように、ゲートタイムはそのノートメッセージの発音時間を示している。なお、図６で示したシーケンスデータチャンクの構造は、PSeq型及びFSeq型におけるシーケンスデータチャンクにおいても同様である。

このシーケンスデータチャンクでサポートされるイベントとしては、次の３通

りのイベントがある。なお、以下に記述する初期値は、イベント指定がないときのデフォルト値である。

（a-1-1）ノートメッセージ「0x9n kk gt」

ここで、ｎ：チャンネル番号（0x0[固定]）、kk：TSeqデータ番号（0x00～0x7F）、gt：ゲートタイム（１～３バイト）である。

ノートメッセージは、チャンネル番号ｎで指定されるチャンネルのTSeqデータ番号kkで指定されるTSeqデータチャンクを解釈し発音を開始するメッセージである。なお、ゲートタイムgtが「0」のノート・メッセージについては発音を行わない。

（a-1-2）ボリューム「0xBn 0x07 vv」

ここで、ｎ：チャンネル番号（0x0[固定]）、vv：コントロール値（0x00～0x7F）である。なお、チャンネルボリュームの初期値は0x64である。

ボリュームは、指定チャンネルの音量を指定するメッセージである。

（a-1-3）パン「0xBn 0x0A vv」

ここで、ｎ：チャンネル番号（0x0[固定]）、ｖｖ：コントロール値（0x00～0x7F）である。なお、パンポット初期値は、0x40（センター）である。

パンメッセージは、指定チャンネルのステレオ音場位置を指定するメッセージである。

【００２３】

（a-2）TSeqデータチャンク（TSeq#00～TSeq#n）

TSeqデータチャンクは、音声合成用の情報として、言語や文字コードに関する情報、発音する音の設定、（合成する）読み情報を表記したテキストなどを含んだ、しゃべり用フォーマットでありタグ形式で書かれている。このTSeqデータチャンクは、ユーザーによる入力を容易にするためテキスト入力となっている。

タグは、”<”（0x3C）で始まり制御タグと値が続く形式であり、TSeqデータチャンクはタグの列で構成されている。ただし、スペースは含まず、制御タグ及び値に”<”は使用することはできない。また、制御タグは必ず１文字とする。制御タグとその有効値に例を下の表４に示す。

【００２４】

【表４】

タグ		値	意味
L	(0x4C)	Language	言語情報
C	(0x43)	*code*	文字ｺｰﾄﾞ名
T	(0x54)	全角文字列	合成用ﾃｷｽﾄ
P	(0x50)	0-	無音の挿入
S	(0x53)	0-127	再生速度
V	(0x56)	0-127	音量
N	(0x4E)	0-127	音の高さ
G	(0x47)	0-127	音色選択
R	(0x52)	None	リセット
Q	(0x51)	None	終了

【００２５】

前記制御タグのうちのテキストタグ「Ｔ」について、さらに説明する。

テキストタグ「Ｔ」に後続する値は、全角ひらがな文字列で記述された読み情報（日本語の場合）と音声表現を指示する韻律記号（Shift-JISコード）からなる。文末にセンテンス区切り記号がないときは、”。”で終わるのと同じ意味とする。

以下に示すのは韻律記号であり、読み情報の文字の後につく。

”、”(0x8141)：センテンスの区切り（通常のイントネーション）。

”。”(0x8142)：センテンスの区切り（通常のイントネーション）。

”？”(0x8148)：センテンスの区切り（疑問のイントネーション）。

”’ ”(0x8166)：ピッチを上げるアクセント（変化後の値はセンテンス区切りまで有効）。

”＿”(0x8151)：ピッチを下げるアクセント（変化後の値はセンテンス区切りまで有効）。

”ー”(0x815B)：長音（直前の語を長く発音する。複数でより長くなる。）

【００２６】

図７の（ａ）は、TSeqデータチャンクのデータの一例を示す図であり、（ｂ）

はその再生時間処理について説明するための図である。

最初のタグ「<LJAPANESE」で言語が日本語であることを示し、「<CS-JIS」で文字コードがシフトＪＩＳであること、「<G4」で音色選択（プログラムチェンジ）、「<V1000」で音量の設定、「<N64」で音の高さを指定している。「<T」は合成用テキストを示し、「<P」はその値により規定されるmsec単位の無音期間の挿入を示している。

図７の（ｂ）に示すように、このTSeqデータチャンクのデータは、デュレーションにより指定されるスタート時点から1000msecの無音期間をおいた後に、「い’やーーー、き＿ょーわ’さ＿むい＿ねー。」と発音され、その後1500msecの無音期間をおいた後に「こ’のままい＿ったら、は’ちが＿つわ、た’いへ’ん＿やねー。」と発音される。ここで、「’」、「＿」、「ー」に応じてそれぞれに対応するアクセントや長音の制御が行われる。

【００２７】

このように、TSeq型は、各国語それぞれに特化した発音をするための文字コードと音声表現（アクセントなど）をタグ形式で記述したフォーマットであるため、エディタなどを用いて直接作成することができる。従って、TSeqデータチャンクのファイルはテキストベースで容易に加工することができ、例えば、記述されている文章からイントネーションを変更したり、語尾を加工することで方言に対応するといったことを容易に行うことができる。また、文章中の特定単語だけを入れ替えることも容易にできる。さらに、データ・サイズが小さいという長所がある。

一方、このTSeq型データチャンクのデータを解釈し音声合成をするための処理負荷が大きくなる、より細かいピッチ制御ができにくい、フォーマットを拡張し複雑な定義を増やせば、ユーザ・フレンドリーでなくなってしまう、言語（文字）コードに依存する（例えば、日本語の場合にはShift-JISが一般であるが、他国語の場合には、それに応じた文字コードでフォーマットを定義する必要がある。）などという短所がある。

【００２８】

（ｂ）PSeq型（フォーマットタイプ＝0x01）

このPSeq型は、ＭＩＤＩイベントに類似する形式の音素によるシーケンス表現（PSeq：phoneme sequence）を用いたフォーマットタイプである。この形式は、音素を記述するようにしているので言語依存がない。音素は発音を示す文字情報により表現することができ、例えば、複数の言語に共通にアスキーコードを用いることができる。

前記図１に示したように、このPSeq型は、セットアップ・データ・チャンク９、ディクショナリ・データ・チャンク１０及びシーケンス・データ・チャンク１１を含んでいる。シーケンスデータ中の音声再生イベント（ノートメッセージ）で指定されたチャンネルの音素と韻律制御情報の再生を指示する。

【００２９】

(b-1) セットアップ・データ・チャンク（Setup Data Chunk）（オプション）

音源部分の音色データなどを格納するチャンクであり、イクスクルーシブ・メッセージの並びを格納する。この実施の形態では、含まれているイクスクルーシブ・メッセージは、ＨＶ音色パラメータ登録メッセージである。

ＨＶ音色パラメータ登録メッセージは「0xF0 Size 0x43 0x79 0x07 0x7F 0x01 PC data ... 0xF7」というフォーマットであり、PC：プログラム番号（0x02～0x0F）、data：ＨＶ音色パラメータである。

このメッセージは、該当するプログラム番号PCのＨＶ音色パラメータを登録する。

【００３０】

ＨＶ音色パラメータを次の表５に示す。

【表５】

#0	基本音声番号
#1	ピッチシフト量[Cent]
#2	フォルマント周波数シフト量１
#3	フォルマント周波数シフト量２
#4	：
#5	フォルマント周波数シフト量ｎ
#6	フォルマントレベルシフト量１
#7	フォルマントレベルシフト量２
#8	：
#9	フォルマントレベルシフト量ｎ
#10	オペレータ波形選択１
#11	オペレータ波形選択２
#12	：
#13	オペレータ波形選択ｎ

【００３１】

表５に示すように、ＨＶ音色パラメータとしては、ピッチシフト量、第１～第ｎ（ｎは２以上の整数）の各フォルマントに対するフォルマント周波数シフト量、フォルマントレベルシフト量及びオペレータ波形選択情報が含まれている。前述のように、処理装置内には、各音素とそれに対応するフォルマント制御情報（フォルマントの周波数、帯域幅、レベルなど）を記述したプリセット辞書（第２の辞書）が記憶されており、ＨＶ音色パラメータは、このプリセット辞書に記憶されているパラメータに対するシフト量を規定している。これにより、全ての音素について同様のシフトが行われ、合成される音声の声質を変化させることができる。

なお、このＨＶ音色パラメータにより、0x02～0x0Fに対応する数（すなわち、プログラム番号の数）の音色を登録することができる。

【００３２】

(b-2) ディクショナリデータチャンク (Dictionary Data Chunk) （オプション）

このチャンクには、言語種別に応じた辞書データ、例えば、前記プリセット辞書と比較した差分データやプリセット辞書で定義していない音素データなどを含

む辞書データを格納する。これにより、音色の異なる個性のある音声を合成することが可能となる。

【００３３】

(b-3) シーケンスデータチャンク (Sequence Data Chunk)

前述のシーケンスデータチャンクと同様に、デュレーションとイベントの組み合わせを時間順に配置したシーケンスデータを含む。

このPSeq型におけるシーケンスデータチャンクでサポートするイベント（メッセージ）を次に列挙する。読み込み側は、これらのメッセージ以外は無視する。また、以下に記述する初期設定値は、イベント指定がないときのデフォルト値である。

【００３４】

(b-3-1) ノートメッセージ「0x9n Nt Vel Gatetime Size data ...」

ここで、n：チャンネル番号（0x0[固定]）、Nt：ノート番号（絶対値ノート指定：0x00～0x7F，相対値ノート指定：0x80～0xFF）、Vel：ベロシティ（0x00～0x7F）、Gatetime：ゲートタイム長（Variable）、Size：データ部のサイズ（可変長）である。

このノートメッセージにより、指定チャンネルの音声の発音が開始される。

なお、ノート番号のＭＳＢは、解釈を絶対値と相対値とに切り替えるフラグである。ＭＳＢ以外の７ビットはノート番号を示す。音声の発音はモノラルのみであるため、ゲートタイムが重なる場合は後着優先として発音する。オーサリング・ツールなどでは、重なりのあるデータは作られないように制限を設けることが望ましい。

【００３５】

データ部は、音素とそれに対する韻律制御情報（ピッチベンド、ボリューム）を含み、次の表６に示すデータ構造からなる。

【表６】

#0	ディレイ
#1	音素数［＝ｎ］
#2	音素１
#3	：
#4	音素ｎ
#5	音素ピッチベンド数［＝Ｎ］
#6	音素ピッチベンド位置１
#7	音素ピッチベンド１
#8	：
#9	音素ピッチベンド位置Ｎ
#10	音素ピッチベンドＮ
#11	音素ボリューム数［＝Ｍ］
#12	音素ボリューム位置１
#13	音素ボリューム１
#14	：
#15	音素ボリューム位置Ｍ
#16	音素ボリュームＭ

【００３６】

表６に示すように、データ部は、音素の数ｎ（#1）、例えばアスキーコードで記述した個々の音素（音素１～音素ｎ）（#2～#4）、及び、韻律制御情報からなっている。韻律制御情報はピッチベンドとボリュームであり、ピッチベンドに関して、その発音区間を音素ピッチベンド数（#5）により規定されるＮ個の区間に区切り、それぞれにおけるピッチベンドを指定するピッチベンド情報（音素ピッチベンド位置１，音素ピッチベンド１（#6～#7）～音素ピッチベンド位置Ｎ，音素ピッチベンドＮ（#9～#10））と、ボリュームに関して、その発音区間を音素ボリューム数（#11）により規定されるＭ個の区間に区切り、それぞれにおけるボリュームを指定するボリューム情報（音素ボリューム位置１，音素ボリューム１（#12,#13）～音素ボリューム位置Ｍ，音素ボリュームＭ（#15,#16））からなっている。

【００３７】

図８は、前記韻律制御情報について説明するための図である。ここでは、発音する文字情報が「ｏｈａｙｏｕ」である場合を例にとって示している。また、こ

の例では、Ｎ＝Ｍ＝１２８としている。この図に示すように、発音する文字情報（「ｏｈａｙｏｕ」）に対応する区間を１２８（＝Ｎ＝Ｍ）の区間に区切り、各点におけるピッチとボリュームを前記ピッチベンド情報及びボリューム情報で表現して韻律を制御するようにしている。

【００３８】

図９は、前記ゲートタイム長（Gatetime）とディレイタイム（Delay Time（#0））との関係を示す図である。この図に示すように、ディレイタイムにより、実際の発音をデュレーションで規定されるタイミングよりも遅らせることができる。なお、Gate time ＝ 0 は、禁止とする。

【００３９】

(b-3-2) プログラムチェンジ「0xCn pp」

ここで、ｎ：チャンネル番号（0x0[固定]）、pp：プログラム番号（0x00～0xFF）である。また、プログラム番号の初期値は0x00とされている。

このプログラムチェンジメッセージにより指定されたチャンネルの音色が設定される。ここで、チャンネル番号は、0x00：男声プリセット音色、0x01：女声プリセット音色、0x02～0x0F：拡張音色である。

【００４０】

(b-3-3) コントロールチェンジ

コントロールチェンジメッセージとしては、次のものがある。

(b-3-3-1) チャンネルボリューム「0xBn 0x07 vv」

ここで、n：チャンネル番号（0x0[固定]）、vv：コントロール値（0x00～0x7F）である。また、チャンネルボリュームの初期値は0x64とされている。

このチャンネルボリュームメッセージは、指定チャンネルの音量を指定するものであり、チャンネル間の音量バランスを設定することを目的としている。

(b-3-3-2) パン「0xBn 0x0A vv」

ここで、n：チャンネル番号（0x0[固定]）、vv：コントロール値（0x00～0x7F）である。パンポットの初期値は0x40（センター）とされている。

このメッセージは、指定チャンネルのステレオ音場位置を指定する。

【００４１】

(b-3-3-3) エクスプレッション「0xBn 0x0B vv」

ここで、n：チャンネル番号（0x0[固定]）、vv：コントロール値（0x00～0x7F）である。このエクスプレッションメッセージの初期値は0x7F（最大値）とされている。

このメッセージは、指定チャンネルのチャンネル・ボリュームで設定した音量の変化を指定する。これは曲中で音量を変化させる目的で使用される。

【００４２】

(b-3-3-4) ピッチベンド「0xEn ll mm」

ここで、n：チャンネル番号（0x0[固定]）、ll：ベンド値ＬＳＢ（0x00～0x7F）、mm：ベンド値ＭＳＢ（0x00～0x7F）である。ピッチベンドの初期値はＭＳＢ0x40、ＬＳＢ0x00とされている。

このメッセージは、指定チャンネルのピッチを上下に変化させる。変化幅（ピッチ・ベンド・レンジ）の初期値は±２半音であり、0x00／0x00で下方向へのピッチ・ベンドが最大となる。0x7F／0x7Fで上方向へのピッチ・ベンドが最大となる。

【００４３】

(b-3-3-5) ピッチベンド・センシティビティ「0x8n bb」

ここで、n：チャンネル番号（0x0[固定]）、bb：データ値（0x00～0x18）である。このピッチベンド・センシティビティの初期値は0x02である。

このメッセージは、指定チャンネルのピッチ・ベンドの感度設定を行う。単位は半音である。例えば、bb＝01のときは±１半音（変化範囲は計２半音）となる。

【００４４】

このように、PSeq型のフォーマットタイプは、発音を示す文字情報で表現した音素単位をベースとし、ＭＩＤＩイベントに類似する形式で音声情報を記述したものであり、データ・サイズはTSeq型よりは大きいがFSeq型よりは小さくなる。

これにより、ＭＩＤＩと同様に時間軸上の細かいピッチやボリュームをコントロールすることができる、音素ベースで記述しているため言語依存性がない、音色（声質）を細かく編集することができる、ＭＩＤＩと類似した制御ができ、従

来のＭＩＤＩ機器へ追加実装し易いという長所を有している。

一方、文章や単語レベルの加工ができない、処理側において、TSeq型よりは軽いものの、フォーマットを解釈し音声合成するための処理負荷がかかるという短所を有している。

【００４５】

（ｃ）フォルマント・フレーム記述（FSeq）型（フォーマットタイプ＝0x02）

フォルマント制御情報（各フォルマントを生成するための、フォルマント周波数やゲインなどのパラメータ）をフレーム・データ列として表現したフォーマットである。すなわち、一定時間（フレーム）の間は、発音する音声のフォルマントなどは一定であるとし、各フレーム毎に発音する音声に対応するフォルマント制御情報（各々のフォルマント周波数やゲインなど）を更新するシーケンス表現（FSeq：formant sequence）を用いる。シーケンスデータに含まれるノートメッセージにより指定されたFSeqデータチャンクのデータの再生を指示する。

このフォーマットタイプは、シーケンスデータチャンクとｎ個（ｎは以上の整数）のFSeqデータチャンク（FSeq#00～FSeq#n）を含んでいる。

【００４６】

（c-1）シーケンスデータチャンク

前述のシーケンスデータチャンクと同様に、デュレーションとイベントの組を時間順に配置したシーケンスデータを含む。

以下に、このシーケンスデータチャンクでサポートするイベント（メッセージ）を列挙する。読み込み側は、これらのメッセージ以外は無視する。また、以下に記述する初期設定値は、イベント指定がないときのデフォルト値である。

（c-1-1）ノート・メッセージ「0x9n kk gt」

ここで、ｎ：チャンネル番号（0x0[固定]）、kk：FSeqデータ番号（0x00～0x7F）、gt：ゲートタイム（1～3バイト）である。

このメッセージは、指定チャンネルのFSeqデータ番号のFSeqデータチャンクを解釈し発音を開始するメッセージである。なお、ゲートタイムが”0”のノート・メッセージは発音を行わない。

【００４７】

（c-1-2）ボリューム「0xBn 0x07 vv」

ここで、n：チャンネル番号（0x0[固定]）、vv：コントロール値（0x00～0x7F）である。なお、チャンネルボリュームの初期値は0x64である。

このメッセージは、指定チャンネルの音量を指定するメッセージである。

【００４８】

（c-1-3）パン「0xBn 0x0A vv」

ここで、n：チャンネル番号（0x0[固定]）、vv：コントロール値（0x00～0x7F）である。なお、パンポットの初期値は0x40（センター）である。

このメッセージは、指定チャンネルのステレオ音場位置を指定するメッセージである。

【００４９】

（c-2）FSeqデータチャンク（FSeq#00～FSeq#n）

FSeqデータチャンクは、FSeqフレーム・データ列で構成する。すなわち、音声情報を所定時間長（例えば、20msec）を有するフレーム毎に切り出し、それぞれのフレーム期間内の音声データを分析して得られたフォルマント制御情報（フォルマント周波数やゲインなど）を、それぞれのフレームの音声データを表わすフレーム・データ列として表現したフォーマットである。

表７にFSeqのフレーム・データ列を示す。

【００５０】

【表７】

#0	オペレータ波形１
#1	オペレータ波形２
#2	：
#3	オペレータ波形ｎ
#4	フォルマントレベル１
#5	フォルマントレベル２
#6	：
#7	フォルマントレベルｎ
#8	フォルマント周波数１
#9	フォルマント周波数２
#10	：
#11	フォルマント周波数ｎ
#12	有声／無声　切り替え

【００５１】

表７において、#0～#3は音声合成に用いる複数個（この実施の形態においては、ｎ個）のフォルマントの波形の種類（サイン波、矩形波など）を指定するデータである。#4～#11は、フォルマントレベル（振幅）（#4～#7）と中心周波数（#8～#11）によりｎ個のフォルマントを規定するパラメータである。#4と#8が第１フォルマント（#0）を規定するパラメータ、以下同様に、#5～#7と#9～#11は第２フォルマント（#1）～第ｎフォルマント（#3）を規定するパラメータである。また、#12は無声／有声を示すフラグなどである。

図１０は、フォルマントのレベルと中心周波数を示す図であり、この実施の形態においては、第１～第ｎフォルマントまでのｎ個のフォルマントのデータを用いるようにしている。前記図４に示したように、各フレーム毎の第１～第ｎフォルマントに関するパラメータとピッチ周波数に関するパラメータは、前記音源部２７のフォルマント生成部とピッチ生成部に供給され、そのフレームの音声合成出力が前述のようにして生成出力される。

【００５２】

図１１は、前記FSeqデータチャンクのボディ部のデータを示す図である。前記表７に示したFSeqのフレームデータ列のうち、#0～#3は、各フォルマントの波形

の種類を指定するデータであり、各フレームごとに指定する必要はない。従って、図１１に示すように、最初のフレームについては、前記表７に示した全てのデータとし、後続するフレームについては、前記表７における#4以降のデータだけでよい。FSeqデータチャンクのボディ部を図１１のようにすることにより、総データ数を少なくすることができる。

【００５３】

このように、FSeq型は、フォルマント制御情報（各々のフォルマント周波数やゲインなど）をフレーム・データ列として表現したフォーマットであるため、FSeq型のファイルをそのまま音源部に出力することにより音声を再生することができる。従って、処理側は音声合成処理の必要がなく、ＣＰＵは所定時間ごとにフレームを更新する処理を行うのみでよい。なお、既に格納されている発音データに対し、一定のオフセットを与えることで音色（声質）を変更することができる。

ただし、FSeq型のデータは文章や単語レベルの加工がしづらく、音色（声質）を細かく編集したり、時間軸上の発音長やフォルマント変位を変更することができない。さらに、時間軸上のピッチやボリュームを制御することはできるが、元のデータのオフセットで制御することとなるため、制御しにくいのに加え、処理負荷が増大するという短所がある。

【００５４】

次に、上述したシーケンスデータのデータ交換フォーマットを有するファイルを利用するシステムについて説明する。

図１２は、上述した音声再生シーケンスデータを再生する音声再生装置の一つである携帯通信端末に対し、上述したデータ交換フォーマットのファイルを配信するコンテンツデータ配信システムの概略構成を示す図である。

この図において、５１は携帯通信端末、５２は基地局、５３は前記複数の基地局を統括する移動交換局、５４は複数の移動交換局を管理するとともに公衆網などの固定網やインターネット５５とのゲートウエイとなる関門局、５６はインターネット５５に接続されたダウンロードセンターのサーバーコンピュータである。

コンテンツデータ制作会社５７は、前記図３に関して説明したように、専用のオーサリング・ツールなどを用い、ＳＭＦやＳＭＡＦなどの楽曲データ及び音声合成用テキストファイルから本発明のデータ交換フォーマットを有するファイルを作成し、サーバーコンピュータ５６に転送する。

サーバーコンピュータ５６には、コンテンツデータ制作会社５７により制作された本発明のデータ交換フォーマットを有するファイル（前記ＨＶトラックチャンクを含むＳＭＡＦファイルなど）が蓄積されており、携帯通信端末５１や図示しないコンピュータなどからアクセスするユーザーからのリクエストに応じて、対応する前記音声再生シーケンスデータを含む楽曲データなどを配信する。

【００５５】

図１３は、音声再生装置の一例である前記携帯通信端末５１の一構成例を示すブロック図である。

この図において、６１はこの装置全体の制御を行う中央処理装置（ＣＰＵ）、６２は各種通信制御プログラムや楽曲再生のためのプログラムなどの制御プログラムおよび各種定数データなどが格納されているＲＯＭ、６３はワークエリアとして使用されるとともに楽曲ファイルや各種アプリケーションプログラムなどを記憶するＲＡＭ、６４は液晶表示装置（ＬＣＤ）などからなる表示部、６５はバイブレータ、６６は複数の操作ボタンなどを有する入力部、６７は変復調部などからなりアンテナ６８に接続される通信部である。

また、６９は、送話マイク及び受話スピーカに接続され、通話のための音声信号の符号化および復号を行う機能を有する音声処理部、７０は前記ＲＡＭ６３などに記憶された楽曲ファイルに基づいて楽曲を再生するとともに、音声を再生して、スピーカ７１に出力する音源部、７２は前記各構成要素間のデータ転送を行うためのバスである。

ユーザーは、前記携帯通信端末５１を用いて、前記図１２に示したダウンロードセンターのサーバー５６にアクセスし、前記３つのフォーマットタイプのうちの所望のタイプの音声再生シーケンスデータを含む本発明のデータ交換フォーマットのファイルをダウンロードして前記ＲＡＭ６３などに格納し、そのまま再生したり、あるいは、着信メロディとして使用することができる。

【００５６】

図１４は、前記サーバーコンピュータ５６からダウンロードして前記ＲＡＭ６３に記憶した本発明のデータ交換フォーマットのファイルを再生する処理の流れを示すフローチャートである。ここでは、ダウンロードしたファイルが、前記図２に示したフォーマットにおいて、スコアトラックチャンクとＨＶトラックチャンクを有するファイルであるとして説明する。

楽曲の再生の開始指示があったとき、或いは、着信メロディとして使用する場合は着信が発生して処理が開始されると、ダウンロードしたファイルに含まれている音声部（ＨＶトラックチャンク）と楽曲部（スコアトラックチャンク）を分離する（ステップＳ１）。そして、音声部については、そのフォーマットタイプが（ａ）TSeq型であるときには、TSeq型をPSeq型に変換する第１のコンバート処理とPSeq型をFSeq型に変換する第２のコンバート処理を実行してFSeq型に変換し、（ｂ）PSeq型であるときには、前記第２のコンバート処理を行ってFSeq型に変換し、（ｃ）FSeq型であるときにはそのままというように、フォーマットタイプに応じた処理を行ってFSeq型のデータに変換し（ステップＳ２）、各フレームのフォルマント制御データをフレーム毎に更新して前記音源部７０に供給する（ステップＳ３）。一方、楽曲部については、音源部に所定のタイミングで楽音発生パラメータを供給する（ステップＳ４）。これにより、音声と楽曲が合成して（ステップＳ５）、出力される（ステップＳ６）。

【００５７】

前記図３に関して説明したように、本発明のデータ交換フォーマットは、ＳＭＦやＳＭＡＦなどの既存の楽曲データ２１に音声合成用テキストデータ２２に基づいて作成した音声再生シーケンスデータを付け加えることにより制作することができるため、上述のように着信メロディなどに利用した場合に多種のエンターテイメント性のあるサービスを提供することが可能となる。

【００５８】

また、上記においてはダウンロードセンターのサーバーコンピュータ５６からダウンロードした音声再生シーケンスデータを再生するものであったが、音声再生装置で上述した本発明のデータ交換フォーマットのファイルを作成することも

できる。

前記携帯通信端末５１において、発声したいテキストに対応する前記TSeq型のTSeqデータチャンクを入力部６６から入力する。例えば、「<Tお’っはよー、げんき？」と入力する。そして、これをそのまま、あるいは、前記第１、第２のコンバート処理を行って、前述の３つのフォーマットタイプのうちのいずれかの音声再生シーケンスデータとし、本発明のデータ交換フォーマットのファイルへ変換して保存する。そして、そのファイルをメールに添付して相手端末に送信する。

このメールを受信した相手方の携帯通信端末では、受信したファイルのタイプを解釈し、対応した処理を行ってその音源部を用いて当該音声を再生する。

このように、携帯通信端末で、データを送信する前に加工することで、多種のエンターテイメント性のあるサービスを提供することが可能となる。この場合、それぞれの加工方法で、サービスに最適な音声合成用フォーマット種類を選択する。

【００５９】

さらにまた、近年では、携帯通信端末においてＪａｖａ(TM)によるアプリケーションプログラムをダウンロードして実行することができるようになっている。そこで、Ｊａｖａ(TM)アプリケーションプログラムを用いてより多彩な処理を行わせることができる。

すなわち、携帯通信端末上で、発声したいテキストを入力する。そして、Ｊａｖａ(TM)アプリケーションプログラムにより、入力されたテキストデータを受け取り、該テキストに合致した画像データ（例えば、しゃべっている顔）を貼付け、本発明のデータ交換フォーマットのファイル（ＨＶトラックチャンクとグラフィックストラックチャンクを有するファイル）へ変換し、Ｊａｖａ(TM)アプリケーションプログラムからＡＰＩ経由で本ファイルをミドルウエア（シーケンサ、音源や画像を制御するソフトウエアモジュール）に送信する。ミドルウエアは送られたファイル・フォーマットを解釈し、音源で音声を再生しながら表示部で画像を同期して表示する。

このように、Ｊａｖａ(TM)アプリケーションのプログラミングにより、多種の

エンターテイメント性のあるサービスを提供することができる。この場合、それぞれの加工方法で、サービスに最適な音声合成用フォーマット種類を選択する。

【００６０】

なお、上述した実施の形態においては、ＨＶトラックチャンクに含まれる音声再生シーケンスデータのフォーマットを３つの型に応じて異なるフォーマットとしていたが、これに限られることはない。例えば、前記図１に示したように、（ａ）TSeq型と（ｃ）FSeq型は、いずれも、シーケンスデータチャンクとTSeqあるいはFSeqデータチャンクを有するものであり、基本的な構造は同一であるので、これらを統一し、データチャンクのレベルで、TSeq型のデータチャンクであるのかFSeq型のデータチャンクであるのかを識別するようにしてもよい。

また、上述した各表に記載したデータの定義は、何れも一例に過ぎないものであり、任意に変更することができる。

【００６１】

【発明の効果】

以上説明したように、本発明の音声再生シーケンスデータのデータ交換フォーマットによれば、音声再生のためのシーケンスを表現することができるとともに、異なるシステムや装置の間で音声再生シーケンスデータを頒布したり交換することが可能となる。

また、楽曲シーケンスデータと音声再生シーケンスデータを各々異なるチャンクに含むようにした本発明のシーケンスデータのデータ交換フォーマットによれば、１つのフォーマット・ファイルで音声再生シーケンスと楽曲シーケンスの同期を取って再生することができる。

また、楽曲シーケンスデータと音声再生シーケンスデータを独立に記述することができ、一方のみを取り出して再生させることが容易にできる。

また、３つのフォーマットタイプを選択することができる本発明のデータ交換フォーマットによれば、音声再生の用途や処理側の負荷を考慮し、最も適切なフォーマットタイプを選択することができる。

【図面の簡単な説明】

【図１】　本発明における音声再生シーケンスデータのデータ交換フォーマ

ットの一実施の形態を示す図である。

【図２】　ＨＶトラックチャンクをデータチャンクの一つとして含むＳＭＡＦファイルの例を示す図である。

【図３】　本発明のデータ交換フォーマットを作成するシステム及び該データ交換フォーマットファイルを利用するシステムの概略構成の一例を示す図である。

【図４】　音源部の概略構成の一例を示す図である。

【図５】　（ａ）TSeq型、（ｂ）PSeq型、及び、（ｃ）FSeq型の３通りのフォーマットタイプの違いについて説明するための図である。

【図６】　（ａ）はシーケンスデータの構成、（ｂ）はデュレーションとゲートタイムの関係を示す図である。

【図７】　（ａ）はTSeqデータチャンクの一例を示す図であり、（ｂ）はその再生時間処理について説明するための図である。

【図８】　韻律制御情報について説明するための図である。

【図９】　ゲートタイムとディレイタイムとの関係を示す図である。

【図１０】　フォルマントのレベルと中心周波数を示す図である。

【図１１】　FSeqデータチャンクのボディ部のデータを示す図である。

【図１２】　音声再生装置の一つである携帯通信端末に対し本発明のデータ交換フォーマットのファイルを配信するコンテンツデータ配信システムの概略構成の一例を示す図である。

【図１３】　携帯通信端末の一構成例を示すブロック図である。

【図１４】　本発明のデータ交換フォーマットのファイルを再生する処理の流れを示すフローチャートである。

【図１５】　ＳＭＡＦの概念を説明するための図である。

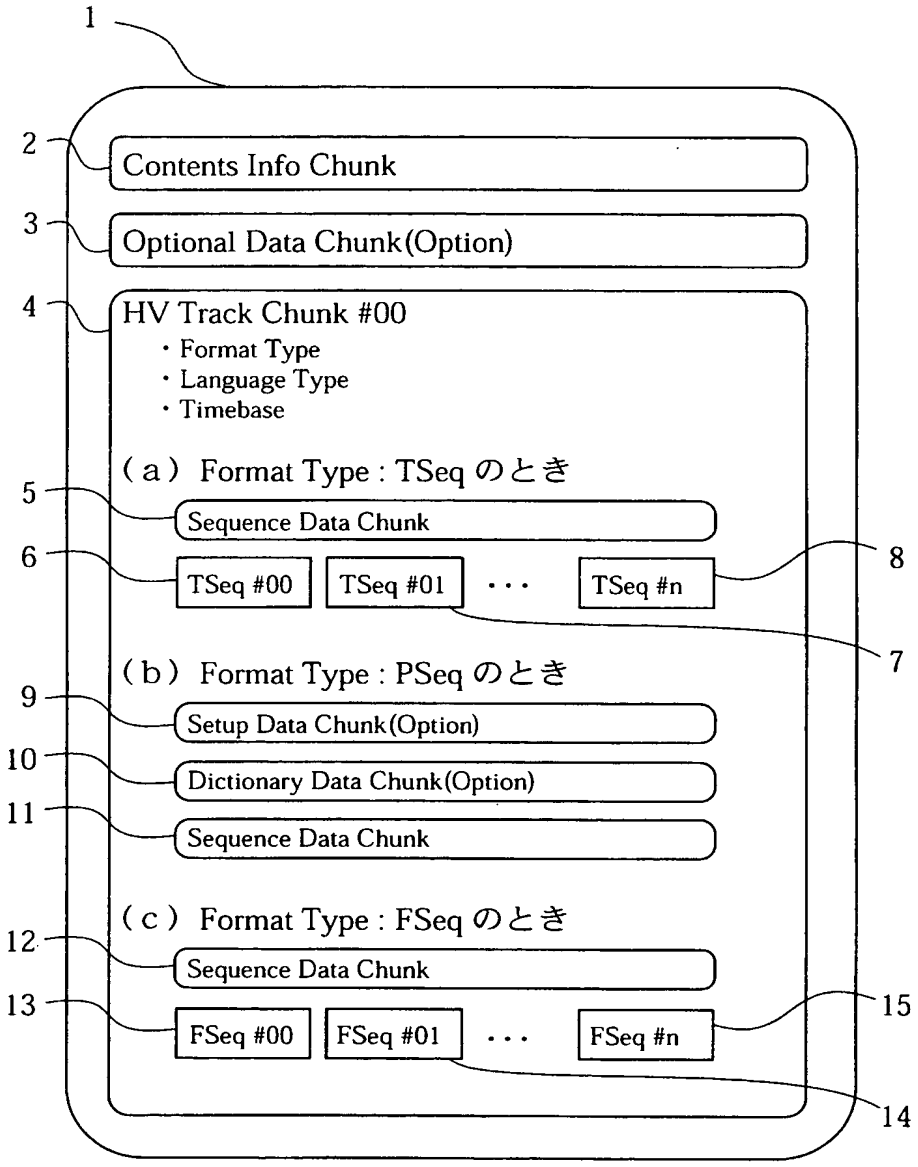
【符号の説明】

１　本発明のデータ交換フォーマットを有するファイル、２　コンテンツ・インフォ・チャンク、３　オプショナル・データ・チャンク、４　ＨＶトラックチャンク、５，１１，１２　シーケンスデータチャンク、６～８　TSeqデータチャンク、９　セットアップデータチャンク、１０　ディクショナリデータチャンク

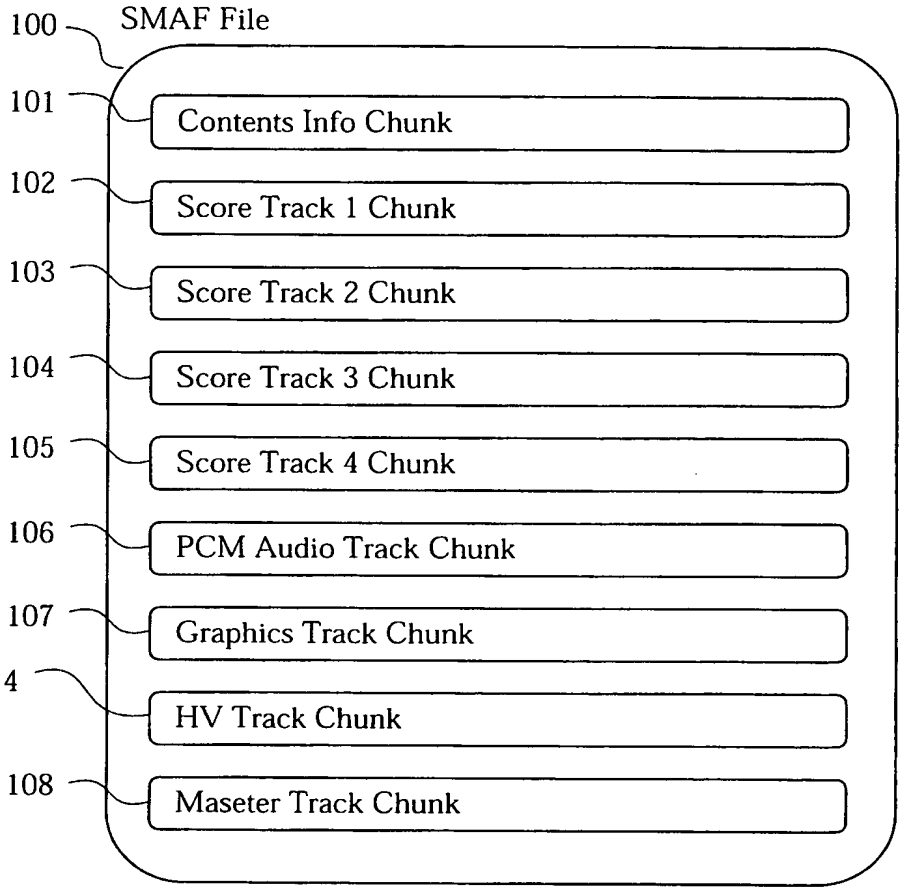
、１３～１５　FSeqデータチャンク、２１　楽曲データ、２２　テキストファイル、２３　オーサリング・ツール、２４　本発明のデータ交換フォーマットを有するファイル、２５　利用装置、２６　シーケンサ、２７　音源部、２８　フォルマント生成部、２９　ピッチ生成部、３０　ミキシング部、５１　携帯通信端末、５２　基地局、５３　移動交換局、５４　関門局、５５　インターネット、５６　ダウンロードサーバー、５７　コンテンツデータ制作会社

【書類名】　図面

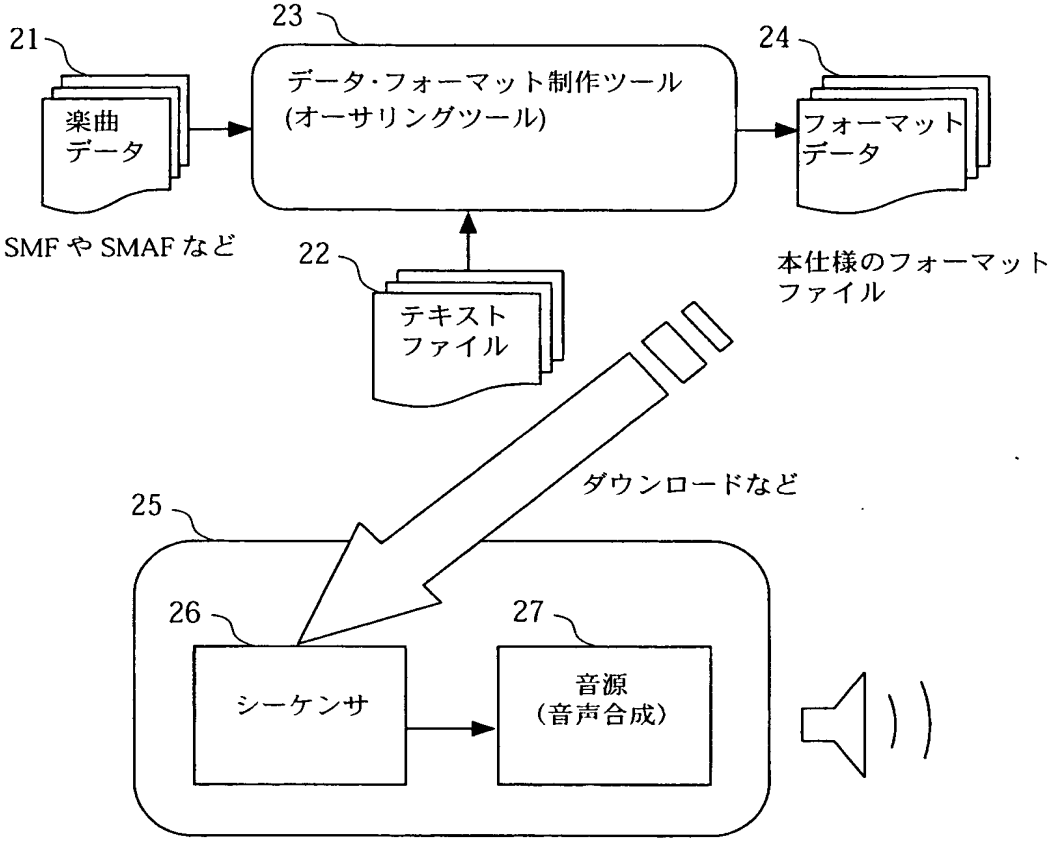
【図１】


1
2
Contents Info Chunk
3
Optional Data Chunk(Option)
4
HV Track Chunk #00
・Format Type
・Language Type
・Timebase
（ａ）Format Type : TSeq のとき
5
Sequence Data Chunk
6
TSeq #00
TSeq #01
・・・
TSeq #n
8
7
（ｂ）Format Type : PSeq のとき
9
Setup Data Chunk(Option)
10
Dictionary Data Chunk(Option)
11
Sequence Data Chunk
（ｃ）Format Type : FSeq のとき
12
Sequence Data Chunk
13
FSeq #00
FSeq #01
・・・
FSeq #n
15
14

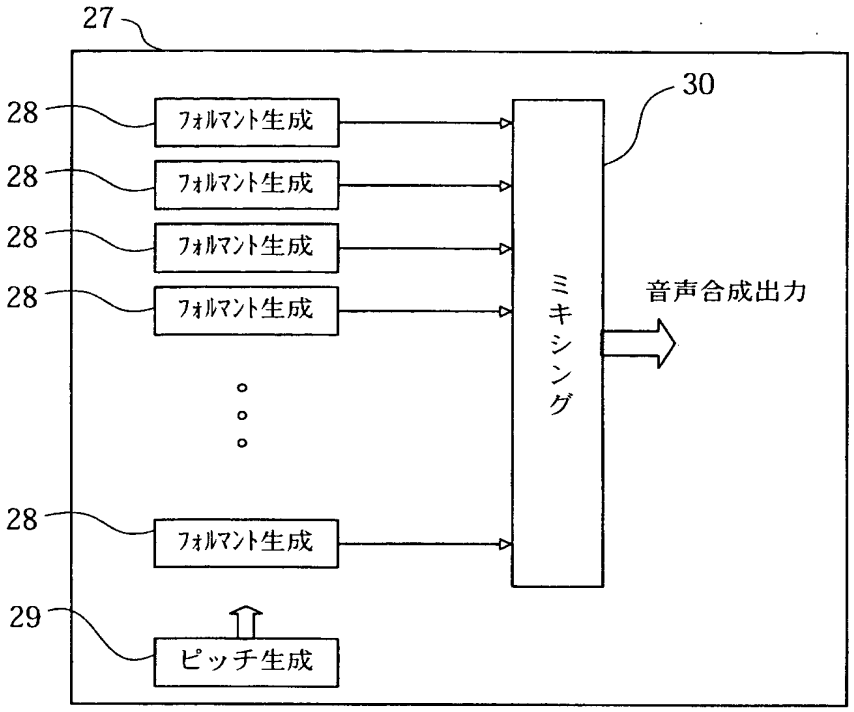
【図２】


100
SMAF File
101
Contents Info Chunk
102
Score Track 1 Chunk
103
Score Track 2 Chunk
104
Score Track 3 Chunk
105
Score Track 4 Chunk
106
PCM Audio Track Chunk
107
Graphics Track Chunk
4
HV Track Chunk
108
Maseter Track Chunk

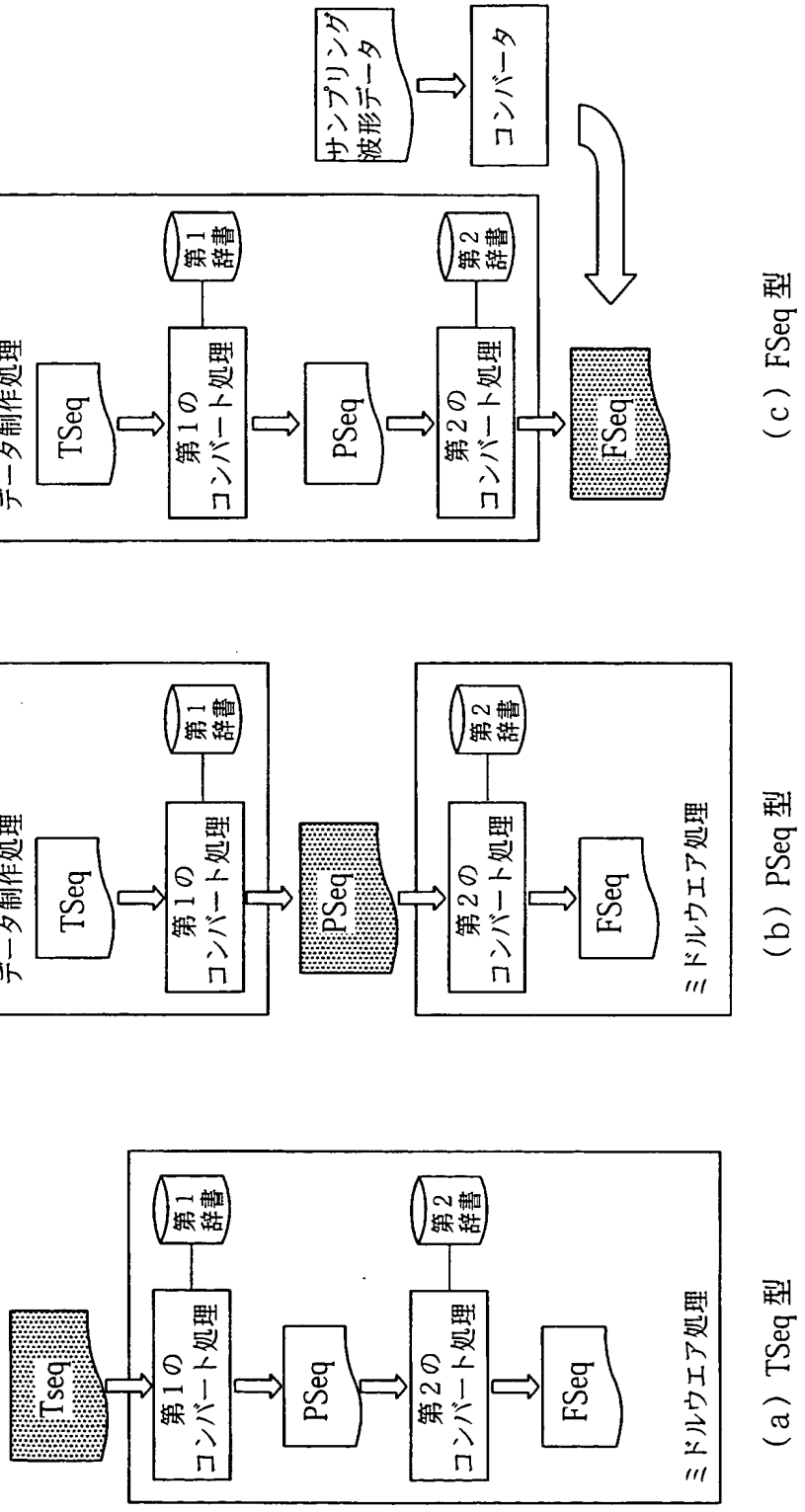
【図３】


21
楽曲
データ
SMF や SMAF など
23
データ･フォーマット制作ツール
(オーサリングツール)
24
フォーマット
データ
本仕様のフォーマット
ファイル
22
テキスト
ファイル
ダウンロードなど
25
26
シーケンサ
27
音源
(音声合成)

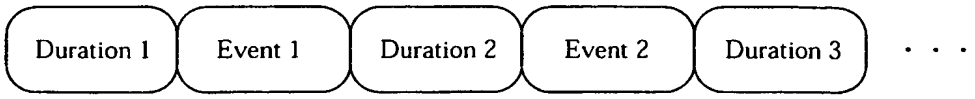
【図４】


27
28
フォルマント生成
28
フォルマント生成
28
フォルマント生成
28
フォルマント生成
28
フォルマント生成
29
ピッチ生成
30
ミキシング
音声合成出力

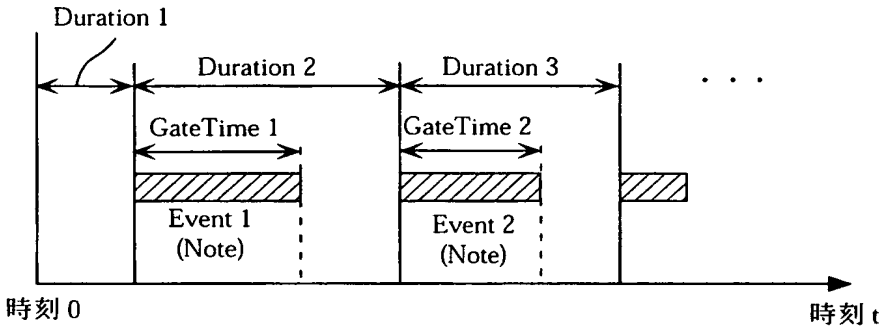
【図５】


Tseq
第１の
コンバート処理
第１
辞書
PSeq
第２の
コンバート処理
第２
辞書
FSeq
ミドルウエア処理
（ａ）TSeq 型
TSeq
第１の
コンバート処理
第１
辞書
PSeq
第２の
コンバート処理
第２
辞書
FSeq
ミドルウエア処理
データ制作処理
（ｂ）PSeq 型
TSeq
第１の
コンバート処理
第１
辞書
PSeq
第２の
コンバート処理
第２
辞書
FSeq
データ制作処理
サンプリング
波形データ
コンバータ
（ｃ）FSeq 型

【図６】


Duration 1
Event 1
Duration 2
Event 2
Duration 3
・・・

（ａ）

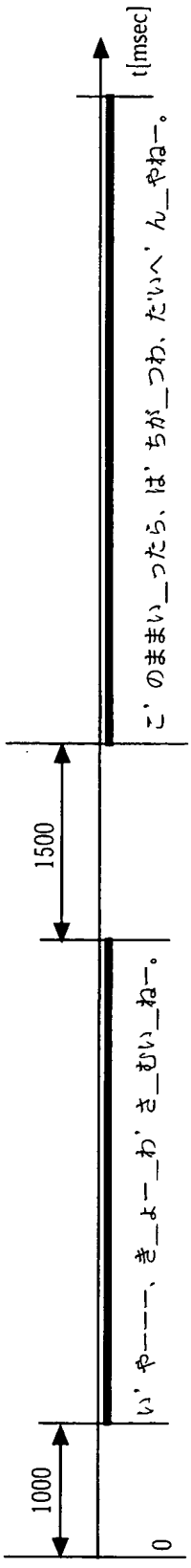

Duration 1
Duration 2
Duration 3
・・・
GateTime 1
GateTime 2
Event 1
(Note)
Event 2
(Note)
時刻 0
時刻 t

（ｂ）

【図７】

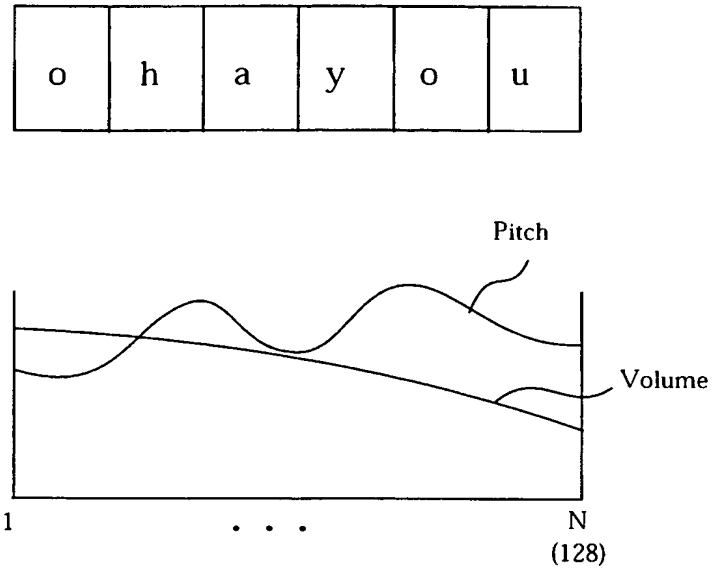
```
<LJAPANESE
<CS-JIS
<G4
<V1000
<N64
<P1000
<Tい'やーーー、き_よー_わ'さ_むい_ねー。
<P1500
<Tこ'のまい_ったら、は'ちが_つわ、たいへ'ん_やねー。
<Q
```

（ａ）TSeq Sample

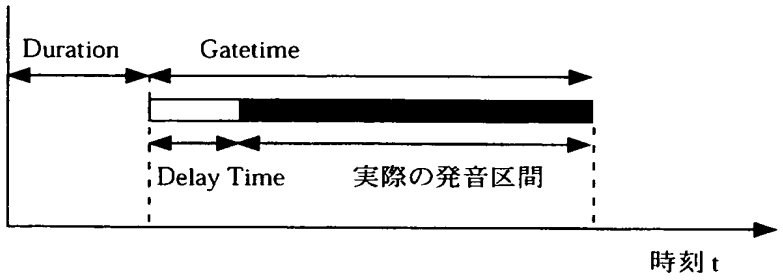


（ｂ）

【図８】

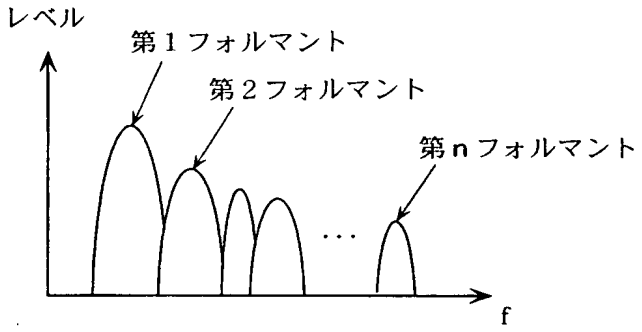


【図９】

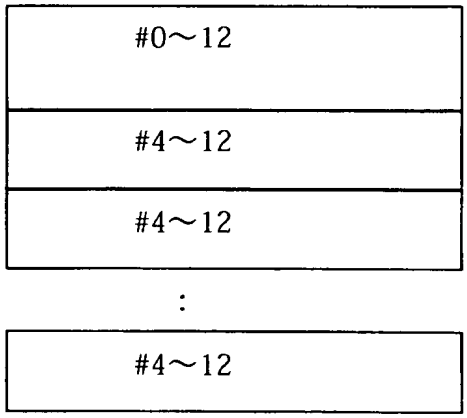


ゲートタイム長とDelay Timeとの関係

【図１０】

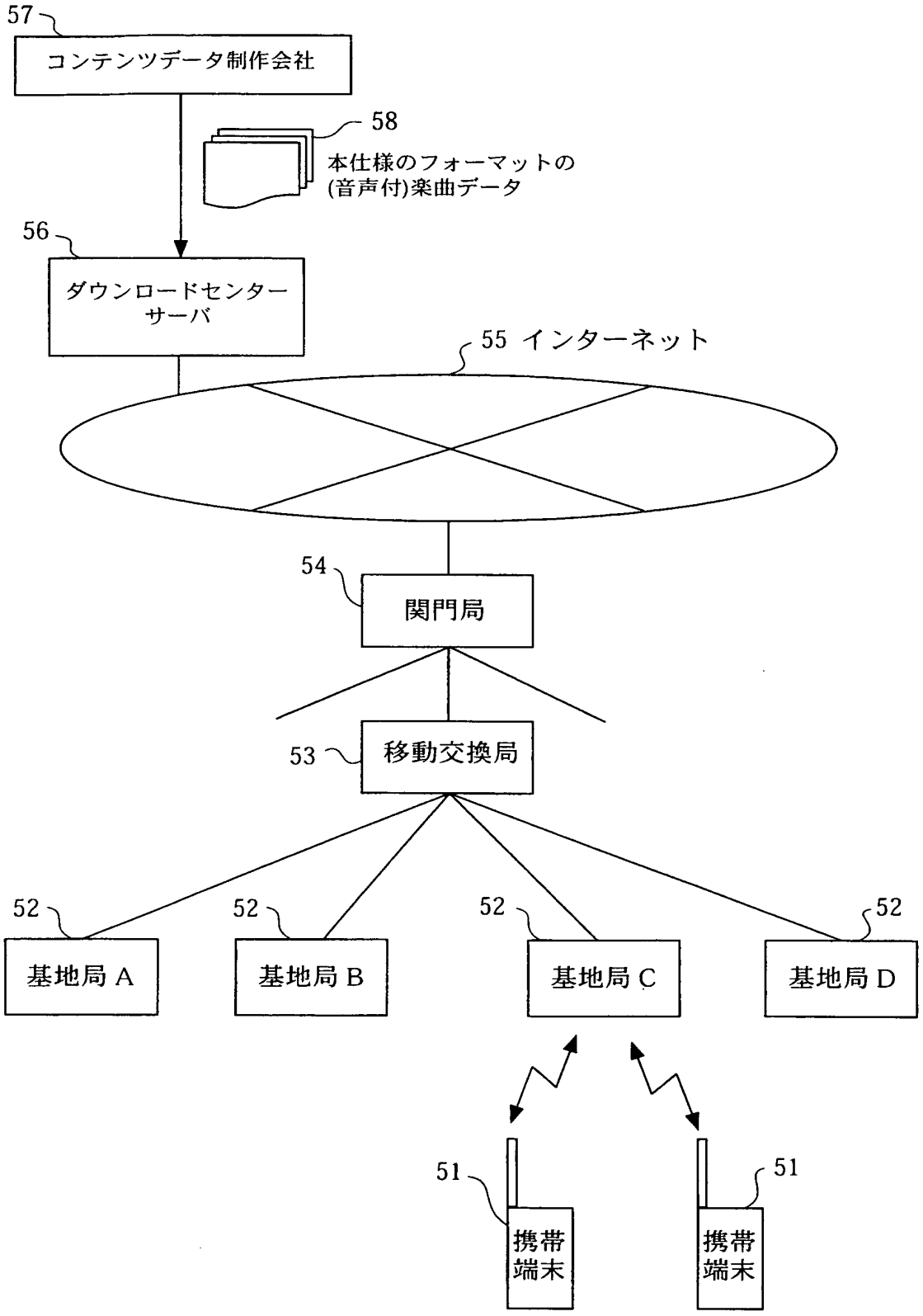


【図１１】

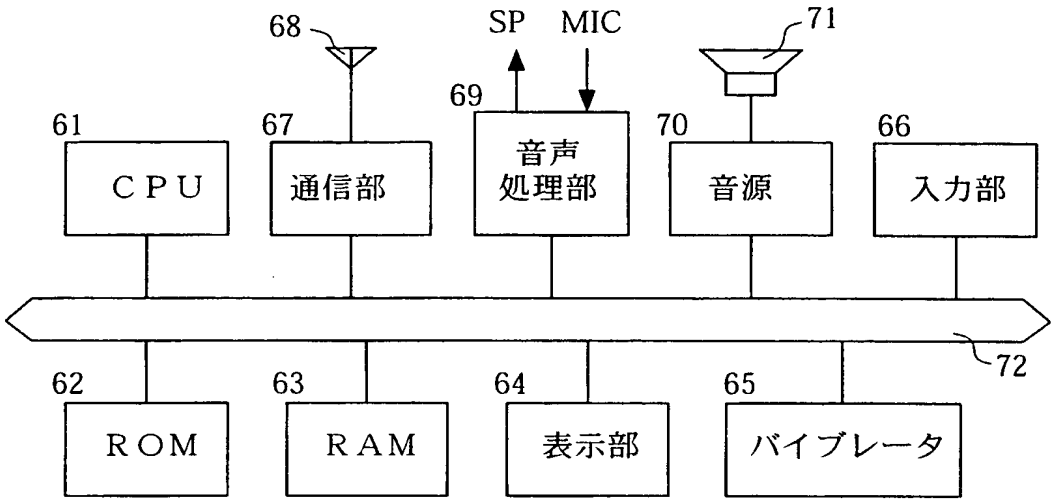


FSeq Data Chunk の Body

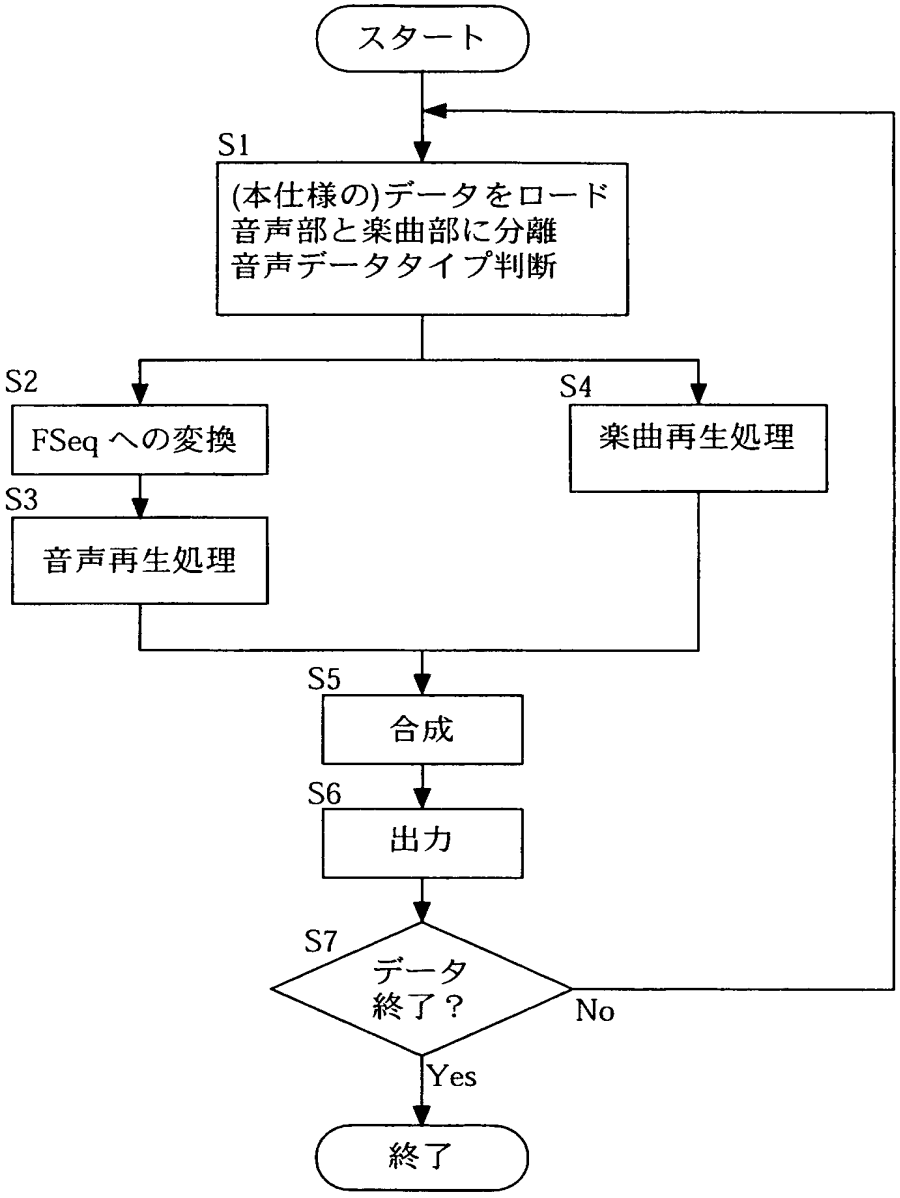
【図１２】


57
コンテンツデータ制作会社
58
本仕様のフォーマットの
(音声付)楽曲データ
56
ダウンロードセンター
サーバ
55 インターネット
54
関門局
53
移動交換局
52
基地局 A
52
基地局 B
52
基地局 C
52
基地局 D
51
携帯
端末
51
携帯
端末

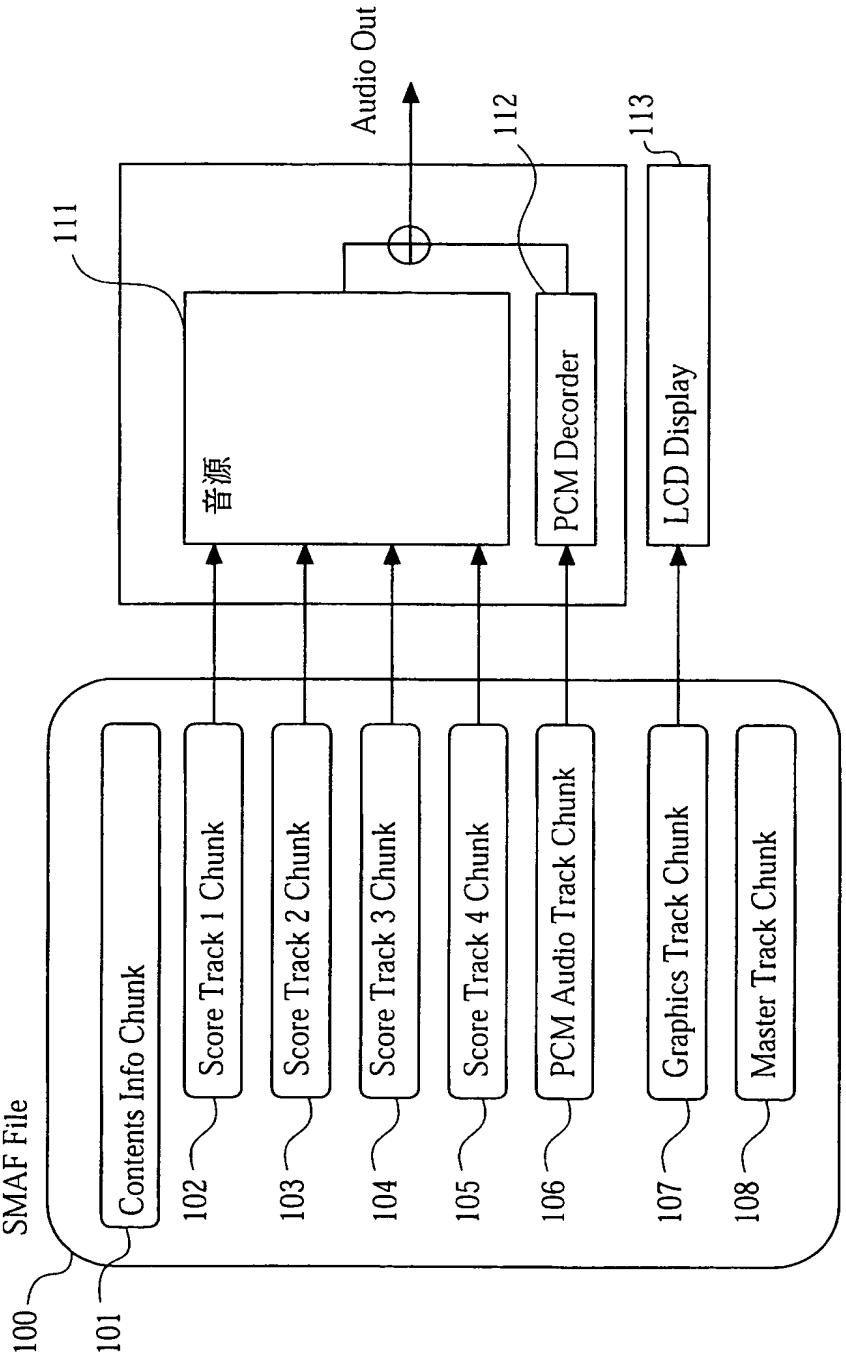
【図１３】


68
SP
MIC
71
69
61
67
70
66
ＣＰＵ
通信部
音声
処理部
音源
入力部
72
62
63
64
65
ＲＯＭ
ＲＡＭ
表示部
バイブレータ

【図１４】


スタート
S1
(本仕様の)データをロード
音声部と楽曲部に分離
音声データタイプ判断
S2
FSeq への変換
S3
音声再生処理
S4
楽曲再生処理
S5
合成
S6
出力
S7
データ
終了？
No
Yes
終了

【図１５】


Audio Out
112
113
111
音源
PCM Decorder
LCD Display
SMAF File
Contents Info Chunk
Score Track 1 Chunk
Score Track 2 Chunk
Score Track 3 Chunk
Score Track 4 Chunk
PCM Audio Track Chunk
Graphics Track Chunk
Master Track Chunk
100
101
102
103
104
105
106
107
108

【書類名】　要約書

【要約】

【課題】　楽曲シーケンスデータと音声再生シーケンスデータを同期して再生することができるシーケンスデータ交換フォーマットを提供する。

【解決手段】　ファイル１はチャンク構造となっており、内部に管理用の情報を含むコンテンツ・インフォ・チャンク２、オプショナル・データ・チャンク３及び音声再生用のＨＶトラックチャンク４を含む。ＨＶトラックチャンク４に含まれる音声再生シーケンスデータは、（１）合成される音声の読みを示すテキスト情報と音声表現を指定する韻律記号とからなるテキスト記述型、（２）合成される音声を示す音素情報と韻律制御情報とからなる音素記述型、又は、（３）再生される音声を示すフレーム時間毎のフォルマント制御情報からなるフォルマントフレーム記述型のいずれかを選択することができる。ＨＶトラックチャンク４は、ＳＭＡＦファイル中にスコアトラックチャンクなどと同様に含ませることができる。

【選択図】　図１

特願２００２－３３５２３３

出　願　人　履　歴　情　報

識別番号　　［０００００４０７５］

１．変更年月日　　１９９０年　８月２２日
［変更理由］　　新規登録
住　所　　静岡県浜松市中沢町１０番１号
氏　名　　ヤマハ株式会社